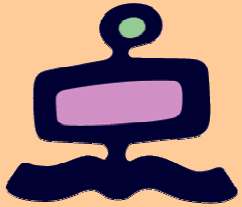


InformaticaUmanistica

Introduzione alle Biblioteche Digitali

UNIVERSITÀ DI PISA

Sommario [1/2]

- **Cenni storici**
  - Vannevar Bush
  - Dalle Biblioteche ai Cataloghi Automatizzati
  - Gli OPAC accessibili via Web
  - Le Biblioteche Digitali
- **Cos'è una Biblioteca Digitale**
  - Definizione
  - Confronto tra BD e database, sistemi IR, WWW, biblioteca tradizionale
  - Vantaggi delle BD
  - Alcuni esempi di Biblioteche Digitali

Sommario [2/2]

- **Cosa ha permesso la nascita delle Biblioteche Digitali**
  - Evoluzioni tecnologiche
- **Tipologie di Biblioteche Digitali**
  - Biblioteche Pubbliche e Biblioteche Specializzate
  - Tipi di documenti trattati
    - ➔ **Libri**
    - ➔ **Documenti testuali**
    - ➔ **Immagini**
    - ➔ **Audio/video**
    - ➔ **……..**

Tipologie di Biblioteche Digitali

Tipologie

- **Biblioteche Pubbliche e Biblioteche Specializzate**
- **Gestione letteratura “white” e “gray”**
- **Gestione di vari tipi di dati**

Biblioteche Pubbliche e Specializzate

- **Una Biblioteca Pubblica prevede che utenti con interessi diversi possano accedere all'informazione**
  - L'accesso è comunque controllato
  - Gli utenti sono costituiti dal grande pubblico
  - I documenti trattano di argomenti diversi
- **Le Biblioteche Specializzate hanno le seguenti caratteristiche**
  - L'insieme degli utenti è piccolo e con interessi molto focalizzati
  - Analogamente i documenti riguardano argomenti molto focalizzati
  - È importante controllare efficacemente l'accesso (utenti non autorizzati, possibilità di visionare gli oggetti ma non di copiarli, ecc.)

Biblioteche Pubbliche e Specializzate

- **Esempi di Biblioteche Specializzate**
  - tradizionali - NASA LaRC Technical Library
  - digitali - NASA Technical Report Server, ACM Digital Library, ETRDL
- **Biblioteche Pubbliche**
  - tradizionali – Biblioteca comunale di Pisa
  - digitali – Yahoo, Boston public library

White and Grey Literature

- **La distinzione tra le due non è sempre molto chiara**
- **La definizione fornita da Grey Net:**
  - "that type of publication unavailable through normal book-selling channels, often produced in small quantities with limited distribution, promotion, and exploitation"
  - http://www.greynet.org/pages/1/index.htm

White and Grey Literature

- **Grey Net ammette comunque che la pubblicazione elettronica ha cambiato questa definizione, che andrà quindi sostituita**
- **Intuitivamente la letteratura**
  - White: autore e publisher sono di solito diversi, il lavoro è stato revisionato in modo indipendente, l'opera può essere ottenuta facilmente
  - Grey: è possibile che non sia stato revisionato; spesso viene pubblicato direttamente dall'autore o dalla sua organizzazione; può essere difficilmente reperibile

Esempi

- **White**
  - Riviste, libri, proceedings di conferenze, etc.
- **Grey**
  - Rapporti tecnici, rapporti governativi, etc.

Gestione di vari tipi di dati nelle BD

- **Libri**
- **Documenti testuali**
- **Immagini**
- **Audio/video**

Gestione di libri

- Automazione accesso ai cataloghi delle Biblioteche tradizionali
- Utilizzo di Cataloghi elettronici
- I servizi della Biblioteca rimangono gli stessi delle Biblioteche tradizionali, la ricerca dei libri risulta più veloce, ed è possibile effettuare ricerche complesse (ad es. libri scritti congiuntamente da due autori, oppure i libri su un certo argomento scritti in un dato periodo, ecc.)
- Si utilizzano tecnologie tradizionali per la gestione dei cataloghi
- È importante uniformare i cataloghi di varie biblioteche per permettere ricerche di libri su più cataloghi

Gestione di documenti testuali

- **Il primo passo dalle Biblioteche tradizionali alle Biblioteche Digitali prevede che la biblioteca abbia i documenti in forma elettronica, non solo i cataloghi**
- **La forma più semplice di contenuto (ma anche quella di più facile utilizzo) è il testo**
- **Documenti testuali ottenuti in modi diversi**
  - Creati direttamente per accesso on-line
  - Convertiti da stampe
  - Digitalizzati dalle tracce audio di film o programmi televisivi

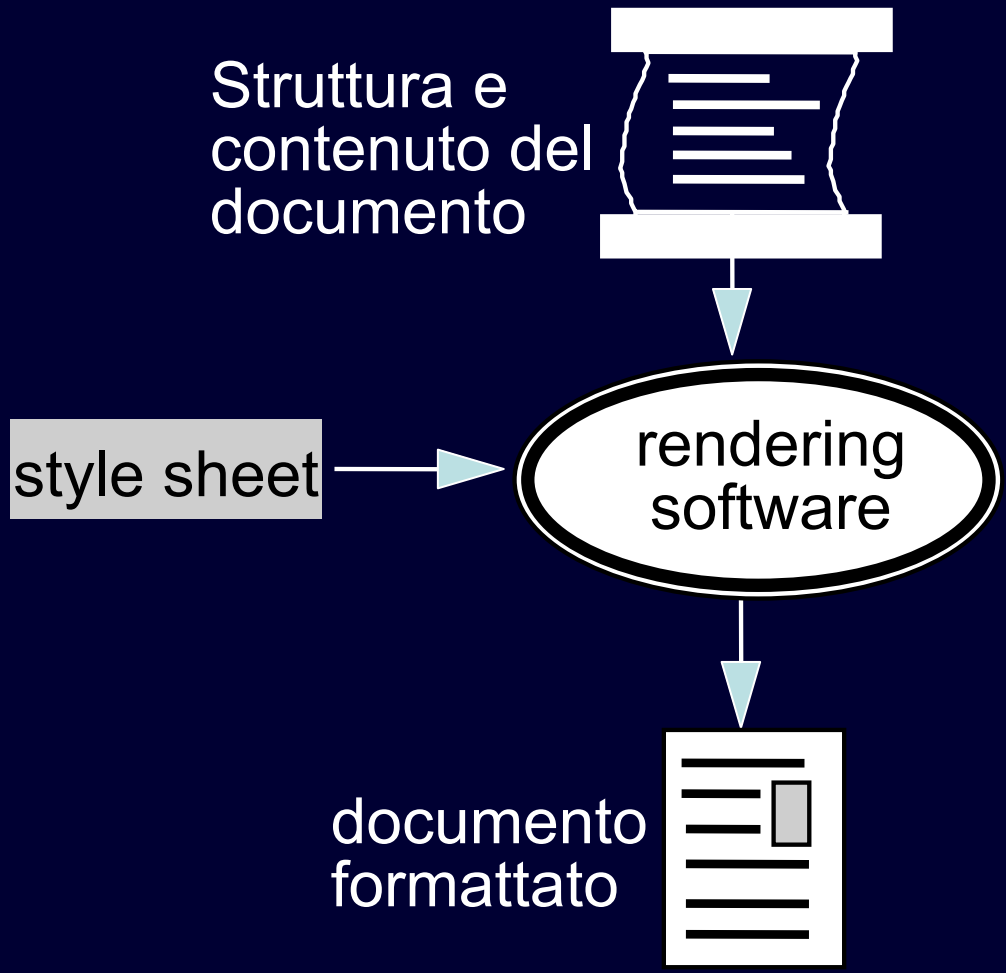
Rappresentazione di documenti testuali

- **Struttura**
  - Descrive la divisione del testo in vari elementi sia fisici (caratteri, parole) che logici (titolo, autori, capitoli, ecc.)
  - La struttura viene spesso rappresentata da linguaggi di markup
- **Linguaggi di Markup**
  - SGML (Standard Generalized Markup Language)
  - HTML/XML
- **Visualizzazione**
  - Descrive il modo in cui il documento viene visualizzato sullo schermo
- **Linguaggi di visualizzazione**
  - TeX, PostScript, PDF

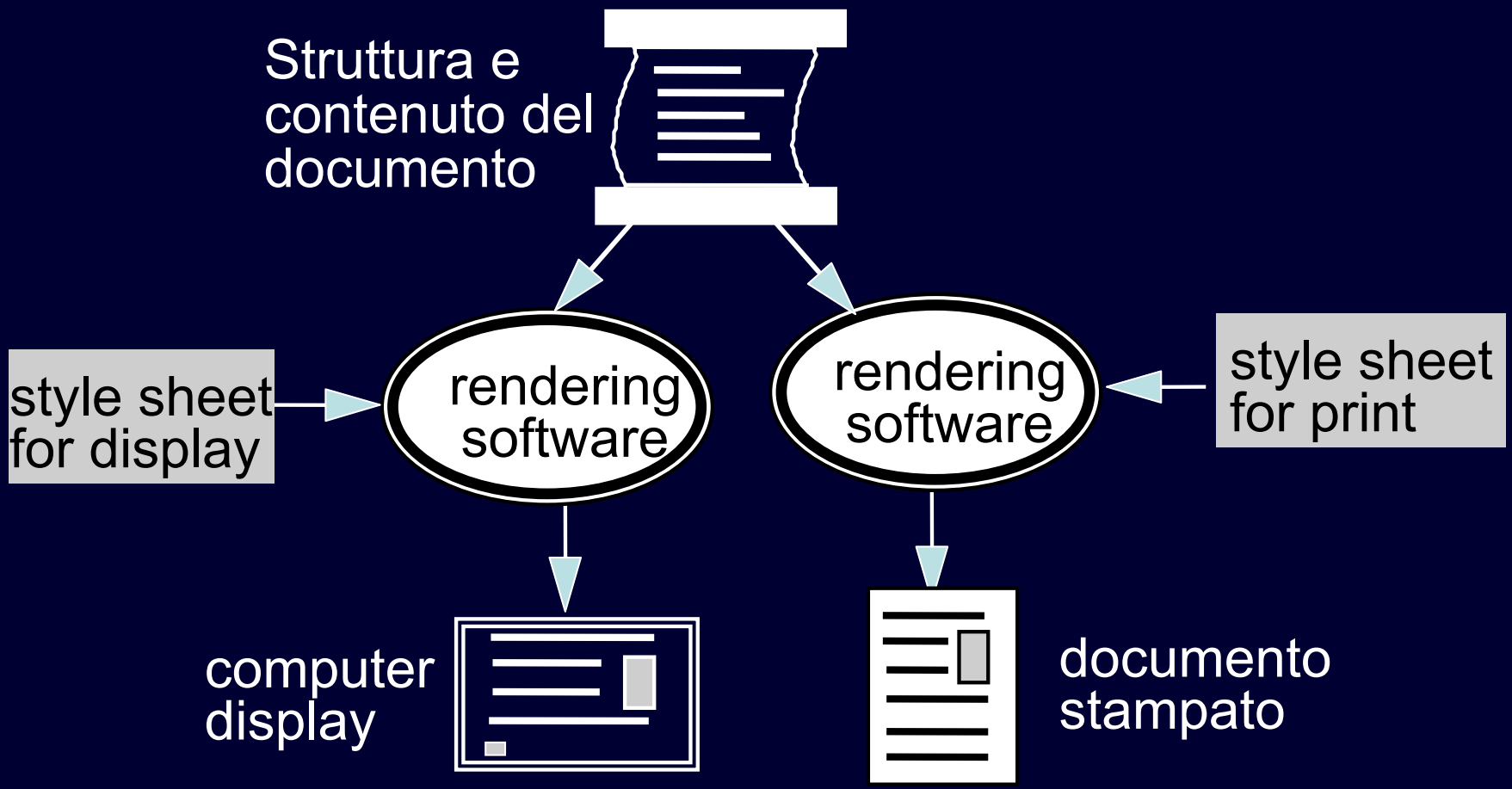
Testo

- **La ricchezza del testo**
  - **Elementi: lettere, scripts, simboli**
  - **Struttura: parole, frasi, paragrafi, titoli, tabelle**
  - **Presentazione: fonts, layout, disegni**
  - **Casi particolari: simboli matematici, musica**
- ***Le Biblioteche Digitali devono rappresentare tutte queste varianti***

Markup e Style Sheets



Alternative Renderings



Markup Languages

SGML (Standard Generalized Markup Language)

A system for creating markup languages that represent the structure of a document

XML (eXtensible Markup Language)

A simplified version of SGML intended for use with online information

DTD (Data Type Definition)

A markup specification for a class of documents, defined within the SGML framework

HTML (Hypertext Markup Language)

A markup and formatting language with links to other objects

XML Example (Metadata)

```
<?xml version="1.0"?>
<!DOCTYPE dlib-meta0.1 SYSTEM "http://www.dlib.org/dlib/dlib-meta01.dtd">
<dlib-meta0.1>
   <title>Digital Libraries and the Problem of Purpose</title>
   <creator>David M. Levy</creator>
  <publisher>Corporation for National Research Initiatives</publisher>
   <date date-type = "publication">January 2000</date>
   <type resource-type = "work">article</type>
```

continued on next slide

XML Example (Metadata)

continued from previous slide

```
<identifier uri-type = "DOI">10.1045/january2000-levy</identifier>
   <identifier uri-type =
"URL">http://www.dlib.org/dlib/january00/01levy.html</identifier>
   <language>English</language>
   <relation rel-type = "InSerial">
      <serial-name>D-Lib Magazine</serial-name>
      <issn>1082-9873</issn>
      <volume>6</volume>
      <issue>1</issue>
   </relation>
   <rights>Copyright (c) David M. Levy</rights>
</dlib-meta0.1>
```

Page-Description Languages

- **Lo scopo è quello di presentare i documenti elettronici con una qualità simile a quella dei documenti a stampa**
- **I primi metodi di formattazione del testo erano specifici per la stampa**
- **Attualmente sono altrettanto importanti le problematiche relative alla visualizzazione su schermo**
- **Vedremo brevemente tre diversi strumenti**
  - TeX – Produzione e formattazione di document
  - PostScript – Stampa di alta qualità
  - Portable Document Format (PDF)

TeX

- Linguaggio sviluppato agli inizi degli anni ’80 da Donald Knuth
- Al contenuto del documento vengono aggiunti una serie di comandi che danno le direttive di formattazione e visualizzazione.
- Contiene istruzioni specializzate per la notazione matematica
- Include un sistema specifico (Metafont) per il disegno di font

PostScript

- Linguaggio grafico sviluppato dalla Adobe Systems, utilizzato principalmente per la creazione di rappresentazioni grafiche di document da stampare
- Molti programmi di gestione documenti possono produrre una rappresentazione PostScript del documento da inviare a device di stampa
- Vi possono essere piccole variazioni dovute ai vari interpreti PostScript
- Utilizzato anche per la memorizzazione e lo scambio di documenti

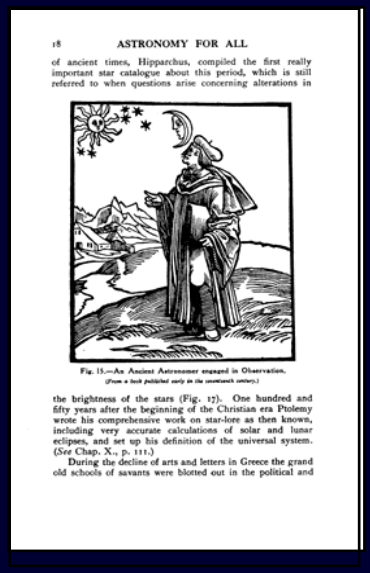
Portable Document Format (PDF)

- Sviluppato dalla Adobe come linguaggio di memorizzazione di pagine di documenti in un formato portabile su diversi sistemi
- Utilizzato principalmente per documenti creati in forma elettronica
- Documenti acquisiti da scanner (bit-map) possono essere estremamente grandi in PDF
- Questo implica che in alcune situazioni il PDF può essere poco adatto ad essere usato nelle Biblioteche Digitali
- I lettori di file PDF sono gratuiti, mentre i programmi di generazione di PDF sono a pagamento

Acquisizione di documenti come immagine

- **Le singole pagine sono acquisite come immagini tramite uno scanner**
- **Ogni singola pagina viene rappresentata come una sequenza di punti (pixels)**
- **Ad ogni pixel viene assegnato un valore (nero, bianco, grigio, colore), rappresentato con un codice binario**
- **Si possono applicare tecniche di compressione della codifica per ridurre l'occupazione dell'immagine**

Metodi di scanning


18 ASTRONOMY FOR ALL

of ancient times, Hipparchus, compiled the first really important star catalogue about this period, which is still referred to when questions arise concerning alterations in

Fig. 15.—An Ancient Astronomer engaged in Observation.

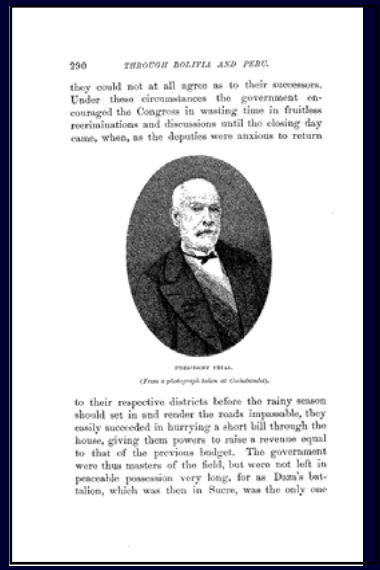
the brightness of the stars (Fig. 17). One hundred and fifty years after the beginning of the Christian era Ptolemy wrote his comprehensive work on star-lore as then known, including very accurate calculations of solar and lunar eclipses, and set up his definition of the universal system. (*See* Chap. X., p. 111.)

During the decline of arts and letters in Greece the grand old schools of savants were blotted out in the political and

Bitonal

Grayscale

Color


290 THROUGH BOLIVIA AND PERU.

they could not at all agree as to their successors. Under these circumstances the government encouraged the Congress in wasting time in fruitless recriminations and discussions until the closing day came, when, as the deputies were anxious to return

to their respective districts before the rainy season should set in and render the roads impassable, they easily succeeded in hurrying a short bill through the house, giving them powers to raise a revenue equal to that of the previous budget. The government were thus masters of the field, but were not left in peaceable possession very long, for as Daza's battalion, which was then in Sucre, was the only one

Special Treatment

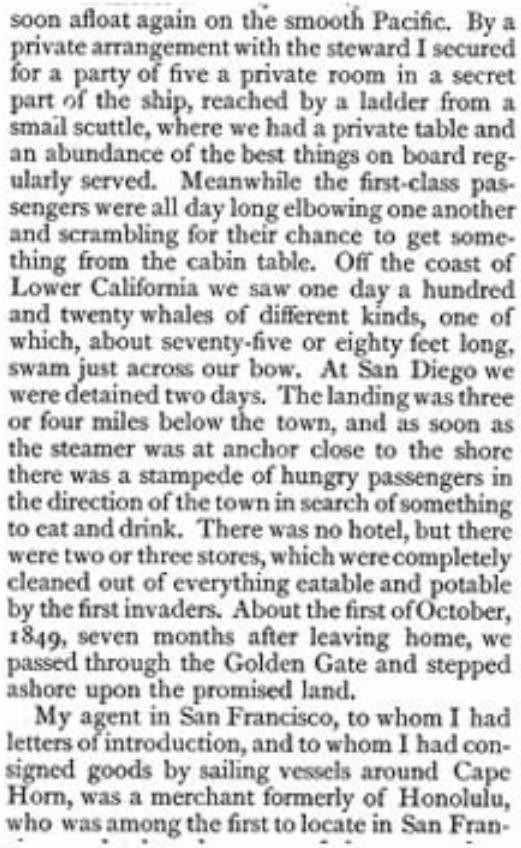
Qualità dell'immagine acquisita

- **Dipende da**
  - Risoluzione dello scanner (numero di pixel per cm^2)
  - bit depth (numero di bit per per pixel)
  - image enhancement
  - color management
  - compression
  - system performance
  - operator judgment and care

Riconoscimento caratteri

- Processo di trasformazione di una immagine in testo
- Dipende dalla qualità dell'immagine
- Utilizza tecniche di image processing combinate con tecniche linguistiche (ad es. utilizzo di dizionari)
- Il risultato è affetto da errori
- In una Biblioteca Digitale è opportuno, in generale, conservare sia l'immagine originale che il testo riconosciuto.
- Il testo può essere utilizzato per la ricerca per contenuto del documento

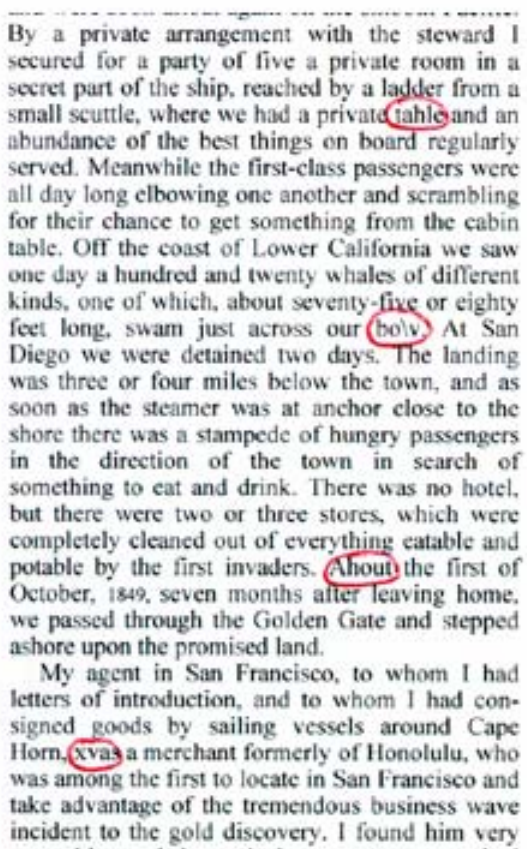
OCR Comparison



soon afloat again on the smooth Pacific. By a private arrangement with the steward I secured for a party of five a private room in a secret part of the ship, reached by a ladder from a smail scuttle, where we had a private table and an abundance of the best things on board regularly served. Meanwhile the first-class passengers were all day long elbowing one another and scrambling for their chance to get something from the cabin table. Off the coast of Lower California we saw one day a hundred and twenty whales of different kinds, one of which, about seventy-five or eighty feet long, swam just across our bow. At San Diego we were detained two days. The landing was three or four miles below the town, and as soon as the steamer was at anchor close to the shore there was a stampede of hungry passengers in the direction of the town in search of something to eat and drink. There was no hotel, but there were two or three stores, which were completely cleaned out of everything eatable and potable by the first invaders. About the first of October, 1849, seven months after leaving home, we passed through the Golden Gate and stepped ashore upon the promised land.

My agent in San Francisco, to whom I had letters of introduction, and to whom I had consigned goods by sailing vessels around Cape Horn, was a merchant formerly of Honolulu, who was among the first to locate in San Fran-

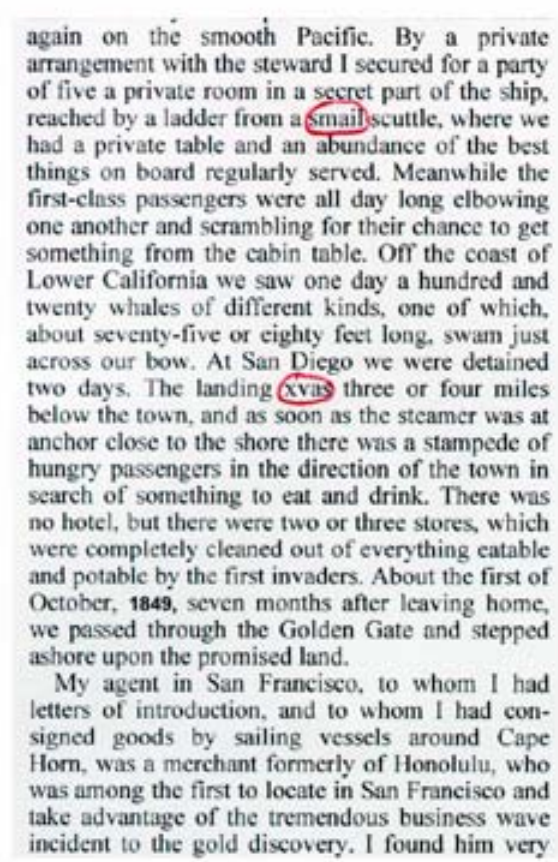
Image of Original Text



By a private arrangement with the steward I secured for a party of five a private room in a secret part of the ship, reached by a ladder from a small scuttle, where we had a private tahle and an abundance of the best things on board regularly served. Meanwhile the first-class passengers were all day long elbowing one another and scrambling for their chance to get something from the cabin table. Off the coast of Lower California we saw one day a hundred and twenty whales of different kinds, one of which, about seventy-five or eighty feet long, swam just across our bo\v. At San Diego we were detained two days. The landing was three or four miles below the town, and as soon as the steamer was at anchor close to the shore there was a stampede of hungry passengers in the direction of the town in search of something to eat and drink. There was no hotel, but there were two or three stores, which were completely cleaned out of everything eatable and potable by the first invaders. Ahout the first of October, 1849, seven months after leaving home, we passed through the Golden Gate and stepped ashore upon the promised land.

My agent in San Francisco, to whom I had letters of introduction, and to whom I had consigned goods by sailing vessels around Cape Horn, xvas a merchant formerly of Honolulu, who was among the first to locate in San Francisco and take advantage of the tremendous business wave incident to the gold discovery. I found him very

OCR Results from 300 dpi Image File



again on the smooth Pacific. By a private arrangement with the steward I secured for a party of five a private room in a secret part of the ship, reached by a ladder from a smail scuttle, where we had a private table and an abundance of the best things on board regularly served. Meanwhile the first-class passengers were all day long elbowing one another and scrambling for their chance to get something from the cabin table. Off the coast of Lower California we saw one day a hundred and twenty whales of different kinds, one of which, about seventy-five or eighty feet long, swam just across our bow. At San Diego we were detained two days. The landing xvas three or four miles below the town, and as soon as the steamer was at anchor close to the shore there was a stampede of hungry passengers in the direction of the town in search of something to eat and drink. There was no hotel, but there were two or three stores, which were completely cleaned out of everything eatable and potable by the first invaders. About the first of October, **1849**, seven months after leaving home, we passed through the Golden Gate and stepped ashore upon the promised land.

My agent in San Francisco, to whom I had letters of introduction, and to whom I had consigned goods by sailing vessels around Cape Horn, was a merchant formerly of Honolulu, who was among the first to locate in San Francisco and take advantage of the tremendous business wave incident to the gold discovery. I found him very

OCR Results from 600 dpi Image File

OCR Comparison of 6 PT Type

Original Image File:

The above reproduction represents the first number of a periodical published by American emigrants belated in Panama in 1849. The original consisted of four pages, about six by ten inches in size, and was printed on light blue paper. So far as we are informed, but four numbers appeared, the others being published on the 3d, 10th, and 17th of March. This publication throws interesting side lights on the Panama trip, of which there is an account. Lists of arrivals are printed in each number. Washington's Birth-

Results of OCR from 600 DPI File:

The above reproduction represents the first number of a periodical published by American emigrants belated in Panama in I34~. The original cousisted of four pages, about six by ten incbes in size, andwas printed 3(8 ligbtblue paper. So far as we are informed, b,tfournumbers appeared, tim others being published onthe 3d, ioth, and ~ytb of March. This publication throws interesting side lights on the P3033333 trip, of which there is an account. Lists of arrivals are printed in each nu3ober. Washington's Birthday, it appears, was duly celebrated by American citizens with a procession and a banquet, and

Results of OCR from 300 DPI File:

Tias above osposslsalisaa osyoehs'slb lbs llsat 'sasatabse of A porisaliassi psbliblatal by Aassoita's sasigosaslla bslAlssl i's Pa'sassd isa 0849. 'Its soigisill hsahiblsal sf fsso psgsa, alassl ais by lsis inasbea in sias, and a'sas psisassal so lighl blass papso. las faa ad ate ate latfoossasal, basl fs'so ossasabass a1spsAssal. alas ohlastb issiog psblislssal s's lts vl~, aslt, asasi sylla of Massla. Tlsib p'sblisalista llaada'sh i'slssshlioag hide higlalb sta llts PsCalsA 10115, of aalaisla llters lb A's asassassl Lib5'sasfaooitdll ats poisled ba sataslastassobst